International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I11P235

Volume 70 Issue 11, 324-337, November 2022

Original Article

Student Employment Context and Learning Achievement Cluster Forecasting Model for Educational Technologists

Tipparat Sittiwong[1], Wongpanya S. Nuankaew[2], Patchara Nasa-Ngium[3], Pratya Nuankaew[4*]

[1] *Faculty of Education, Naresuan University, Phitsanulok, Thailand.*
[2] *Faculty Information Technology, Rajabhat Maha Sarakham University, Maha Sarakham, Thailand.*
[3] *Faculty Science and Technology, Rajabhat Maha Sarakham University, Maha Sarakham, Thailand.*
[4] *School of Information and Communication Technology, University of Phayao, Phayao, Thailand.*

[4*]*Corresponding Author : pratya.nu@up.ac.th*

Received: 22 August 2022 Revised: 10 November 2022 Accepted: 17 November 2022 Published: 26 November 2022

Abstract *- This research investigated the insight analysis of learning patterns that affect graduates' employment prospects. There are three objectives: 1) to study the context and learning patterns that encourage learners' achievement through educational data mining techniques, 2) to develop the forecasting models and construct an application for forecasting appropriate academic achievement clusters, and 3) to assess the application for forecasting the appropriate academic achievement clusters. Samples were collected from 227 students at the Department of Educational Technology and Communication, the Faculty of Education, Naresuan University, during the academic year 2015 – 2020. Research instruments were divided into two categories: statistical and data mining. The results showed that the context of the students in the curriculum offered a high level of academic performance distributed at high levels of academic achievement, which was 3.58 of an average overall student GPA. In addition, the developed forecasting model provided a high level of accuracy, which was 71.11 percent accuracy. At the same time, the level of satisfaction with the application was the highest level of satisfaction, which was 4.21 of the overall satisfaction with 0.66 of S.D. It can be concluded that this research is valuable and beneficial to educational technologists, who are driving the adoption of technology in education.*

Keywords *- Insight Analysis, Learning Patterns, Educational Data Mining, Forecasting Model.*

1. Introduction

Nowadays, the situation of education in Thailand is entering Education 4.0. It turned the Education 2.0 era, where users relied on basic audiovisual equipment and materials, into the era where learners and instructors fully adopted computer-based technology and instructional support programs. Learners can access learning resources independently when the network system and the Internet are expanded. Moreover, analog technology has been replaced by digital technology since it entered the 3rd era of education or Education 3.0. In 2020, the world's education entered a new phase of Education 4.0. The educational system changed the visions and attitudes of learners and educators by focusing on the mass of knowledge available and accessible globally. Besides, it creates new knowledge and develops innovations to meet the needs of society.

Era transition circumstance emerged an enormous amount of research on Education 4.0 challenges [1]–[5]. Wittayasin [5] indicated the problems that reflected the education system of Thailand as well as the inequality in education. For example, school size does matter, poor English language skills, lack of 21st-century skills, and so on.

It is consistent with the research work of Bunwirat and Boonsathorn [1] that examined factors that affected the education system of Thailand. It comprises 21st-century education and Education 4.0.

Moreover, Siririn [2] showed some aspects of improving the quality of education in higher education. It was suggested that it is necessary to raise the incorporation of latent missions in the form of the application for either innovation technology or information and communication technology in teaching and learning. These problems encountered in the research related to the development of the education system of Thailand towards Education 4.0. The quality of education seriously impacts the student employment and job prospects of new graduates. The characteristics of universities for graduate employment have been discussed under the research variances [6], [7]. It contains variables used to establish the university's reputation and ranking. It also includes the university's academic reputation, employer reputation, international research network, citations per paper, etc. There is little research on the context of educational curricula toward the labor market [8]. This aspect reflects the learner's attitude towards teachers, educational institutions, and completed courses [9]. Moreover, the researchers have not yet found the deployed artificial intelligence and machine learning technologies dedicated to analytics and research in terms of predicting the relationship between educational curriculum and employment among new graduates.

Therefore, the researchers aim to study and develop student employment forecasting models influenced by the educational curriculum. There are three main research objectives. The first objective is to study the context and learning patterns to engage learning through educational data mining techniques of students at the Department of Educational Technology and Communication, the Faculty of Education, Naresuan University. The second objective is to develop a forecasting model of student employment influenced by an educational curriculum with data mining techniques. Finally, the third objective is to construct a student employment forecasting application influenced by an educational curriculum. Furthermore, this research work is based on user satisfaction.

Populations and samples were purposive samples from 227 students at the Department of Educational Technology and Communication, the Faculty of Education, Naresuan University, who were enrolled and succeeded in academic achievement. Research instruments are divided into two parts. The first part of the tool is intended to describe and contextualize students' achievement in the Bachelor of Art Program in Educational Technology and Communications at the Faculty of Education, Naresuan University. It consists of essential statistical tools, including frequency determination, percentage, mean, standard deviation, and interpretation. In addition, it was applied to evaluate user satisfaction with the application. The second part involves using data mining tools and processes to develop models. It consists of a data mining development process with CRISP-DM and five prediction techniques of machine learning tools: decision tree, random forest, gradient boosted trees, decision stump, and random tree. The performance testing tools adopted cross-validation methods and the confusion matrix performance.

Researchers have great confidence that this research will drive and promote success in the education system of Thailand. Moreover, this research will be a model and motivation for the further development of artificial intelligence and machine learning in the Thai education industry.

2. Literature Reviews

These literature reviews summarized the relationship and relevance of modern trending technologies in the educational industry application. In addition, the perspective of employment of educational technologists was given as follows:

2.1. Behavioral Performance Analytics for Student Model

Research work for the study of analyzing and defining learner behavior is a new area. The application of artificial intelligence and machine learning technologies analyzes and synthesizes learners' behavior until discoverable the patterns of success or achievement of learners at all levels [4], [10]–[12]. Moreover, the linkage of big data analytics technology in the education industry is becoming more attractive, known as "Educational Data Mining" [13]–[16]. Many researchers explored student learning styles combined with innovations in data analytics. They studied several theories related to educational development, such as self-regulated learning [13], [14] to support learning activities, clustering and predicting learner success under learning behavior [17], and studying the success factors of learners during the COVID-19 pandemic situation [18].

However, the perspectives and understandings of educational technologists are limited. There are large frames perceived in the social sciences. Most of their research studies aimed to create an understanding of learners and teachers in educational institutions. They lack the technology to directly analyze learner behavior and needs that lead to success and sustainable learning in the future. Therefore, it is imperative to develop science and educational innovation technology.

2.2. Student Employment Forecasts

Concerns about getting a job among students or new graduates are intensifying. Many researchers focused on this issue [19]–[23]. It includes getting a job during the internship [21], obtaining a job for a specific occupation [20], [22], [23], and the impact of cross-country work migration [24]. It seems like student employment is of broad interest.Regarding educational technologists, some graduates do not work in educational technology-related fields, though they achieved high academic achievement. Nonetheless, many researchers aimed to study developments in the area of educational technology and educational technologists [25]–[28]. Undoubtedly, a strong focus on study and research in educational technologists is emerging in the new area where educators' attention is dedicated worldwide. These are the motivations and drives for this research.

3. Materials and Methods

The concept and context of global society are poised for the transformation of learning in the 21st century. The researchers strongly believe that developing technologies which are consistent and conducive to learning behaviours can appropriately promote and support sustainable lifelong learning. Therefore, this research design and control research instruments are divided into four main areas: population and sample determination, terminology definition, research tool development, and research evaluation.

3.1. Population and Sample Determination

The population was students enrolled in the Bachelor of Art Program in Educational Technology and Communications at the Faculty of Education, Naresuan University, during the academic year 2015 – 2020.

The research samples were purposive sampling. There are 227 students at the Department of Educational Technology and Communication, the Faculty of Education, Naresuan University, as summarized in Table 1.

Table 1. Research Samples

Academic Year	Number of Students	Grade Point Average (G.P.A.)				
		Mean	Mode	Median	Min	Max
2015	27 (11.84%)	3.11	3.19	3.13	2.23	3.61
2016	37 (16.23%)	3.09	3.05	3.15	2.21	3.62
2017	34 (14.91%)	3.14	2.86	3.14	2.47	3.67
2018	42 (18.42%)	2.98	3.47	2.89	2.28	3.58
2019	39 (17.11%)	2.82	2.64	2.95	1.86	3.73
2020	49 (21.46%)	3.16	3.03	3.25	1.80	3.80
Total:	**227 (100%)**	**3.05**	**2.86**	**3.12**	**1.80**	**3.80**

Table 1 summarizes the data collection. Samples were collected during the academic year 2015 – 2020. It consisted of 248,969 transactions in the registration system of Naresuan University. Each transaction consists of five attributes: student code, course code, course name, course credits, and student grades.

The collected student data was divided into eight grades as follows. Grade A is equal to the value of 4.00, which means Excellent. Grade B+ is equal to the value of 3.50, which means Very Good. Grade B is equal to the value of 3.00, which means Good. Grade C+ is equal to the value of 2.50, which means Fairly Good. Grade C is equal to the value of 2.00, which means Fair. Grade D+ is equal to the value of 1.50, which means Poor. Grade D is equal to the value of 1.00, which means Very Poor. Grade F is equal to the value of 0, which means Fail. Students must reach at least a 2.00-grade point average for graduation.

According to Table 1, the overall data collection scenario revealed that the learners had a good level of achievement. It has an average academic achievement of the sample at 3.05. In contrast, the mode value showed that most students had grade point averages at a fairly good level (2.86).

3.2. Terminology Definition

To encourage understanding in this section, the explanation of the terminology consists of the following:

3.2.1. Insight Analysis

Insight analysis is an analysis design for understanding the studied data, consisting of data design planning, data collection, data analysis, and summary of analysis results. It came up with informed recommendations and recommended actions. The benefits are trend analysis results, data type clustering results, and heuristic results varied depending on the pattern of each data set. The study of insight analysis can be applied to sustainable modeling and strategic planning.

3.2.2. Learning Patterns

Learning patterns are concepts that connect all learning activities that occur over time to explain learning beliefs and motivations. Learning patterns retain the relationship between learning activities, learning plans, learning assessment methods, and learning assessment outcomes. In addition, learning patterns are also an essential component of learning style inventories. They are qualified to promote decision-making among learners to adapt to the appropriate learning styles.

3.2.3. Educational Data Mining

Educational Data Mining: EDM is an application of data analysis technology based on data mining principles that focuses on finding knowledge or patterns in educational data. It can be used to describe the data analysis process and apply it to the learner development process through formal education to create and develop students to achieve academic achievement.

In addition, educational data mining is also a part of understanding learners' learning behaviors, searching for courses that influence learners' achievement, discovering appropriate learning patterns for individuals, and promoting learning for the development of pertinent learning. Moreover, educational data mining can use its findings to develop educational software to develop learning management systems and design sustainable alternatives for learners.

3.2.4. Forecasting Model

The forecasting model is a model for predicting student achievement based on a learner's learning behavior. It can classify and group learners by using data science methods and machine learning processes to select the right group of learners. Moreover, it can link learner behavior and learning achievement without human bias.

3.3. Research Tool Development

The selected tools are divided into two main areas: basic statistical analysis tools and data mining model development tools.

3.3.1. Basic Statistical Analysis Tools

The purpose of essential statistical analysis tools is to describe the context of the data. In addition, it is used for assessing the satisfaction of the application usage. It consists of the frequency value, percentage, mean, standard deviation, and interpretation.

In terms of rating scales and interpretations, details are described in the research evaluation section.

3.3.2. Data Mining Model Development Tools

The purpose of data mining is to illustrate the process of model development by researchers selected from Cross-Industry Standard Process for Data Mining (CRISP-DM) principles [29], [30]. It is an industry-proven way to guide data mining efforts that contain six steps: business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

Moreover, the researchers strictly followed the CRISP-DM data mining development process in the Business Understanding section. The researchers explored the need of those involved in the Bachelor of Art Program in Educational Technology and Communications at the Faculty of Education, Naresuan University, to determine the problems the course managers would like to solve. Afterwards, they will bring various problems to the design of the data collection. There is a combination of functions in terms of data understanding and data preparation. Researchers coordinated the request to use courtesy of 248,969 transaction data from Naresuan University. It needs to be prepared through the process of data cleaning and transformation for model development.

Regarding the model development process and modeling, the researchers selected five prediction techniques: decision tree, random forest, gradient boosted trees, decision stump, and random tree. These five techniques are popular machine-learning tools for different modeling forecasts. The decision tree is a concrete and easy-to-understand technique, while the random forest technique is known as ensemble learning. This model is based on the same model training multiple times (multiple instances) on the same data set. Besides, the gradient-boosted trees (GBT) technique is an altered decision tree-based technique that optimizes model performance by randomly generating hundreds of decision trees. It evaluates each model until a complete decision tree is obtained. Also, the decision stump technique is a machine-learning model with a single-level decision structure. It is a decision tree from the root node connected to the leaf node. Thus, the decision stump predictions are based on the value of a single input attribute.

Furthermore, the random tree technique works on the same principle as the decision tree technique except for splitting the subset into a single random subset of attributes if necessary. In evaluating model efficiency or evaluation, the researchers applied cross-validation techniques and confusion matrix performance to find the model's effectiveness, as described in the research evaluation section. Finally, in the deployment process, the researchers implemented the most accurate model to develop an application as a tool to summarize and present the solution to the course managers. Moreover, as demonstrated in the evaluation section, this application was tested using a questionnaire to assess system satisfaction.

3.4. Research Evaluation

The research evaluation was divided into two parts:

3.4.1. Model Performance Evaluation

Model performance evaluation is intended to be a tool for selecting effective and suitable models that should be developed into applications. There are two evaluation tools, including cross-validation techniques and confusion matrix performance.

The cross-validation techniques are data segmentation techniques for testing a developed model that divides the data into two datasets. The first dataset is intended to develop a model called the "Training Dataset". The test dataset is used to test the developed model, called the "Testing Dataset". While the confusion matrix is an instrument used to show model performance measurements. It consists of accuracy, precision, and recall [17]. Accuracy is forecast values from developed and tested models with prepared data. Precision is the value that the forecast is based on in the forecast section. Recall is the accuracy of a forecast in a precise answer class. All three elements are adapted altogether to determine and select the most efficient and suitable model.

3.4.2. Application Satisfaction Assessment

The application satisfaction assessment is intended to serve as a tool for reflecting the attitude and satisfaction of the application that has been designed and developed. The assessment was based on a questionnaire with five satisfaction levels of the Likert rating scale principle. It consisted of 5 means very satisfied, 4 means somewhat satisfied, 3 means neither satisfied nor dissatisfied, 2 means somewhat dissatisfied, and 1 mean very dissatisfied.

Interpretation is calculated as the total score divided by the number of sector rates. It consists of five levels as follows: 4.21 – 5.00 means strongly agree, 3.41 – 4.20 means agreeing, 2.61 – 3.40 means neither agree nor disagree, 1.81 – 2.60 means disagree, and 1.00 – 1.80 for strongly disagree. The proposed questions in the questionnaire are shown in Table 2.

The questions in Table 2 were performed in the target group. There were 30 experienced and knowledgeable students in the field of information technology from the Faculty of Education, Naresuan University. The results of the assessment are presented in Table 14.

Table 2. Questions in the Questionnaire

Section/ State	Questions
Assessment of Functionality and Procedures	
State 1	Satisfaction with the steps and processes of the system
State 2	Satisfaction with the features and functionality of the system
Assessment of the System Performance	
State 3	Satisfaction with the accuracy and precision of the system
State 4	Satisfaction with the objectives required by the system
State 5	Satisfaction with the user-friendly design and simple menu
State 6	Satisfaction with the modernity of the information
Assessment of Convenience and UX-UI Design	
State 7	Satisfaction with user experience design
State 8	Satisfaction with user interaction design
Assessment of the System Quality	
State 9	Satisfaction with the overall system usage
State 10	Satisfaction with the system's capabilities and utilization

4. Research Results

The research results were divided into three sections that correspond to the research objectives: the context of curriculum productivity, modelling and application development, and satisfaction with the application. The details are presented as follows:

4.1. Context of Curriculum Productivity

The essence of this section is to find out the context and the learning model for students at the Department of Educational Technology and Communication, Faculty of Education, Naresuan University. Moreover, the researchers would like to present the key factor in student achievement.

The sample data presented in Table 1 were classified into two categories. The 1st category was the 96 students who had already completed an educational program, and the 2nd category was the 131 students who were studying in an educational program. After the students graduated from Naresuan University, surveys were conducted where the students who graduated will provide information for the student's employment as summarized in Table 3.

Please note that the study program takes four years to manage to learn, so learners who started in 2015 graduated in 2018 and so forth.

Table 3 shows the number and percentage of respondents to the employment survey for the academic year 2019 – 2021. The respondents graduated in the academic year 2018 – 2020. There are four key indicators: Idt^1 shows the number of employed graduates. Idt^2 shows the number of unemployed graduates. Idt^3 shows the number of graduates being employed directly with the educational program. Idt^4 shows the number of employed graduates but does not match the educational program.

The survey responses were at a high level of cooperation, with all three years having a total of 87 respondents representing 96.67 percent. In addition, the number of graduates employed by 57 respondents was 63.33 percent, as shown in Indicator 1 (Idt^1). However, there were only 37 respondents who were employed directly in the education program representing 41.11 percent, as shown in Indicator 3 (Idt^3).

It preliminarily concludes that educational programs should expedite the search and improvement of educational programs to meet the demands of the future labor market. In addition, there was a conflict when considering the consistency of learners who attended and graduated, as concluded in Table 4.

Table 4 presents data and statistics on graduation from educational programs. According to the data, there are approximately 8.16% of late graduating students. Even if the students graduated according to the study plan, the proportions of being employed were opposite. Table 1 found that the overall grade point average was very high (Mean = 3.05).

It was supported by a high graduation rate average of 91.84 percent, as shown in Table 4. However, student employment was only 63.33%, and 41.11% matched the educational program goal, as shown in Table 3.

Therefore, the context for producing graduates in educational technology contradicts the labor market.

4.2. Modeling and Application Development

This section consists of four steps: determining student achievement levels, developing an achievement forecasting model, testing the effectiveness of the achievement model, and developing an application as described below.

Table 3. The Students' Employment

Year	Graduate	Respondents	Indicators			
			Idt1	Idt2	Idt3	Idt4
2019	25	24 (96.00%)	21 (84.00%)	3 (12.00%)	11 (44.00%)	10 (40.00%)
2020	35	35 (100.00%)	14 (40.00%)	21 (60.00%)	12 (34.29%)	2 (5.71%)
2021	30	28 (93.33%)	22 (73.33%)	6 (20.00%)	14 (46.67%)	8 (26.67%)
Total	**90**	**87 (96.67%)**	**57 (63.33%)**	**30 (33.33%)**	**37 (41.11%)**	**20 (22.22%)**

Table 4. Conflict of Achievement in Educational Program

Generation	Students Attended	Students Graduated	Conflict of Achievement
2015 – 2018	27 (27.55%)	25 (25.51%)	-2 (-2.04%)
2016 – 2019	37 (37.76%)	35 (35.71%)	-2 (-2.04%)
2017 – 2020	34 (34.69%)	30 (30.61%)	-4(-4.08%)
Total:	**98 (100%)**	**90 (91.84%)**	**-8 (-8.16%)**

4.2.1. Determining Student Achievement Levels

The main goal of this section is to investigate the relationship between academic achievement before enrolling in an educational program defined as "Admission GPA" after enrolling in an educational program. It is defined as "Graduation GPA", as summarized in the data for analysis in Table 5. This section's outcome is a cluster representing the student's achievement level. Table 5 shows that the overall learners in the educational program had higher academic achievement after graduation. This data section was analyzed to determine the learner's level and achievement cluster.

Table 5. Academic Achievement of Students

Enrollment Year	Grade Point Average (G.P.A.)				
	Mean	Mode	Median	Min	Max
Admission GPA					
2015	2.98	1.83	2.99	1.83	3.78
2016	2.87	2.55	2.89	2.04	3.55
2017	2.76	2.63	2.84	1.98	3.48
Total:	**2.87**	**2.69**	**2.89**	**1.83**	**3.78**
Graduation GPA					
2015	3.12	3.19	3.13	2.23	3.61
2016	3.12	3.05	3.17	2.21	3.62
2017	3.16	2.86	3.16	2.72	3.67
Total:	**3.13**	**2.86**	**3.16**	**2.21**	**3.67**

Table 6. Average within Centroid Distance Value

Number of Clusters	AVD	Number of Clusters	AVD	Number of Clusters	AVD
2	0.1332	9	0.0326	16	0.0166
3	0.0969	10	0.0277	17	0.0147
4	0.0750	11	0.0231	18	0.0144
5	0.0610	12	0.0246	19	0.0148
6	0.0531	13	0.0203	20	0.0135
7	0.0417	14	0.0198		
8	0.0368	15	0.0186		

The student's achievement level was determined using data mining techniques using two clustering techniques, K-Means and K-Medoids. Both techniques are important tools in segmentation techniques. In addition to determining the appropriate cluster value, the k-determination technique is used for consideration [31]. The results of the analysis are presented in Figure 1 to Figure 2 and Table 6 to Table 11.

Figure 1 demonstrates a cluster analysis with the k-Means technique of the relationship between academic achievement before and after enrolling in an educational program. It found that the optimal cluster number from the analysis of the k-value was equal to 7. In addition, the average within centroid distance (AVD) and member distributions for each cluster were presented in Table 6 and Table 7. Additionally, the centroid values of each cluster are presented in Table 8.

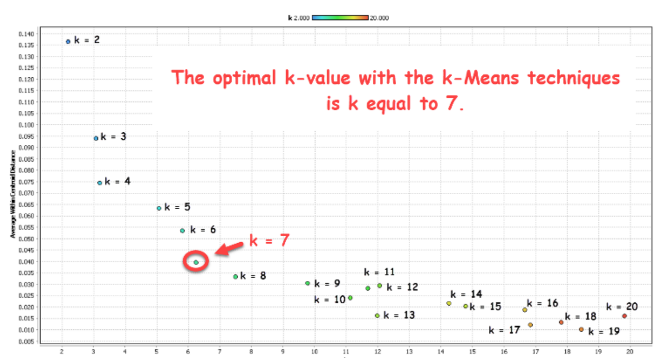

Fig. 1 Optimal k-value with k-Means technique

Table 7. Member Distributions

Cluster	Number of Member	Cluster	Number of Member
Cluster 0	4 items	Cluster 4	23 items
Cluster 1	11 items	Cluster 5	10 items
Cluster 2	22 items	Cluster 6	13 items
Cluster 3	7 items		

* Total number of items: 90

Table 8. Centroid Values of Each Cluster

Attribute	Cluster			
	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Admission GPA	2.96	3.42	2.72	2.65
Graduation GPA	3.23	3.49	2.64	3.42
	Cluster 4	Cluster 5	Cluster 6	
Admission GPA	2.41	3.29	1.98	
Graduation GPA	3.08	2.99	2.59	

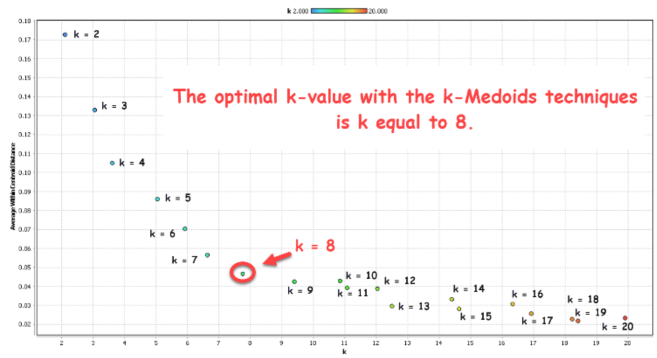

Fig. 2 Optimal k-value with k-Medoids technique

Table 9. Average within Centroid Distance Value

Number of Clusters	AVD	Number of Clusters	AVD	Number of Clusters	AVD
2	0.1707	9	0.0439	16	0.0294
3	0.1348	10	0.0427	17	0.0273
4	0.1054	11	0.0388	18	0.0254
5	0.0846	12	0.0359	19	0.0228
6	0.0702	13	0.0322	20	0.0218
7	0.0579	14	0.0330		
8	0.0488	15	0.0263		

Table 10. Member Distributions

Cluster	Number of Member	Cluster	Number of Member
Cluster 0	21 items	Cluster 4	11 items
Cluster 1	7 items	Cluster 5	8 items
Cluster 2	15 items	Cluster 6	14 items
Cluster 3	4 items	Cluster 7	10 items

* Total number of items: 90

From Table 6 to Table 8, the k-Means clustering analysis shows details of various considerations. It will be used for comparison with the k-Medoids technique, where the comparative results will be used to create a clustered forecast of student achievement.

Figure 2 presents a cluster analysis with the k-Medoids technique of the relationship between academic achievement before and after enrolling in an educational program. It found that the optimal cluster number from the analysis of the k-value was equal to 8. In addition, the average within centroid distance (AVD) and member distributions for each cluster were presented in Table 9 and Table 10. Additionally, the centroid values of each cluster are presented in Table 11.

Table 11. Centroid Values of Each Cluster

Attribute	Cluster			
	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Admission GPA	2.63	3.48	3.00	2.63
Graduation GPA	3.09	2.83	3.34	3.67
	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Admission GPA	3.42	2.69	3.03	2.08
Graduation GPA	3.64	2.86	3.08	2.86

From Table 9 to Table 11, the k-Medoids clustering analysis shows details of various considerations. It will be used for comparison with the k-Means technique, where the comparative results will be used to create a clustered forecast of student achievement.

The following process brings the appropriate cluster to create a forecast model, which is presented in the next section.

4.2.2. Developing An Achievement Forecasting Model

This section aims to develop a forecasting model that can predict the reasonable clustering of clusters to plan learner support in an educational program. The tools used in this section are forecasting techniques, also known as "Supervised Learning" based on machine learning techniques. The techniques used in this study include decision trees, random forests, gradient-boosted trees, decision stumps, and random trees.

The data used to create the forecasting model was the student's GPA admission and grades in the first academic year, consisting of 13 courses. The cluster data (class) was obtained from the cluster clustering model using the k-Means technique and the k-Medoids technique. Therefore, the model derived from this research consists of two significant models for academic achievement. The model shows the outcome of the relationship between the student's GPA admission grades and both two-type of course groups.

Table 12. Model Analysis Results with Cluster Data from k-Means

Modeling Techniques	Accuracy with CV Performance		
	10-Fold	50-Fold	Leave-One-Out
• Decision Tree	67.78%	64.44%	65.56%
• Random Forest	52.22%	61.00%	60.00%
• Gradient Boosted Trees*	71.11%*	65.00%	66.67%
• Decision Stump	37.78%	35.56%	34.44%
• Random Tree	40.00%	35.56%	26.67%

Model testing and selection with various techniques were preliminarily considered by cross-validation techniques and confusion matrix performance, as reported in Table 12 and Table 13.

Table 12 presents the results of the analysis to develop a forecasting model for clustering student achievement by using five analytical techniques with data from k-Means clustering. It found that the gradient-boosted trees (GBT) technique provided the highest accuracy of 71.11%. The detailed model analysis is shown in Figure 3.

Table 13. Model Analysis Results with Cluster Data from k-Medoids

Modeling Techniques	Accuracy with CV Performance		
	10-Fold	50-Fold	Leave-One-Out
Decision Tree*	67.78%	67.00%	70.00%*
Random Forest	61.11%	55.56%	55.56%
Gradient Boosted Trees	66.67%	64.44%	66.67%
Decision Stump	38.89%	35.56%	34.44%
Random Tree	34.44%	35.00%	23.33%

Table 13 presents the analysis results to develop a forecasting model for clustering student achievement using five analytical techniques with data from k-Medoids clustering. It found that the decision tree technique offered the highest accuracy of 70.00%. The detailed model analysis is shown in Figure 4.

The summary results from Table 13 and Table 14 are presented in the following sections of the model performance analysis.

4.2.3. Testing the Effectiveness of The Achievement Model

This section presents a model performance analysis divided into two parts. The first part is the data analysis of the model from the cluster using the k-Means technique, and the second part is the data analysis of the model from the cluster using the k-Medoids technique. The results of the analysis are shown in Figures 3 and Figure 4.

According to the model performance analysis, compared between Table 13 to Table 14 and Figure 3 to Figure 4, the researchers decided to use a data clustering model of the k-Means clustering technique that organized 7 suitable clusters. Nevertheless, the model efficiency (accuracy value) was 71.11% from the gradient-boosted trees (GBT) technique. Finally, the model was developed into an application, as presented in the next section.

4.2.4. Developing An Application

This section presents applications that have been developed from the learner achievement cluster forecasting model. It is used to devise strategies to support teaching and learning. The chosen models were those that derived from the decision tree technique, as shown in the analysis data in Table 13 and Figure 4. The application that has been developed and tested is presented in Figures 5 to Figures 8.

	true cluster_0	true cluster_1	true cluster_2	true cluster_3	true cluster_4	true cluster_5	true cluster_6	class precision
pred.cluster_0	0	0	0	0	1	0	0	0.00%
pred.cluster_1	0	7	7	0	0	0	2	77.78%
pred.cluster_2	0	0	18	0	0	3	0	85.71%
pred.cluster_3	3	0	0	7	1	0	0	63.64%
pred.cluster_4	1	0	0	0	19	4	0	79.17%
pred.cluster_5	0	0	2	0	2	3	1	37.50%
pred.cluster_6	0	4	2	0	0	0	10	62.50%
class recal	0.00%	63.64%	81.82%	100.00%	82.61%	30.00%	76.92%	

Fig. 3 Model Performance Testing with k-Means Data

	true cluster_0	true cluster_1	true cluster_2	true cluster_3	true cluster_4	true cluster_5	true cluster_6	true cluster_7	class precision
pred.cluster_0	9	9	0	0	0	6	0	1	56.25%
pred.cluster_1	0	2	0	0	1	0	0	0	66.67%
pred.cluster_2	0	0	20	5	1	0	1	0	74.07%
pred.cluster_3	0	0	0	1	0	0	0	2	33.33%
pred.cluster_4	0	1	1	0	5	0	0	0	71.43%
pred.cluster_5	1	0	0	0	0	0	0	1	0.00%
pred.cluster_6	0	0	0	0	0	1	2	0	66.67%
pred.cluster_7	1	0	0	2	0	0	2	24	82.76%
class recal	81.82%	66.67%	95.24%	12.50%	71.43%	0.00%	40.00%	85.71%	

Fig. 4 Model Performance Testing with k-Medoids Data

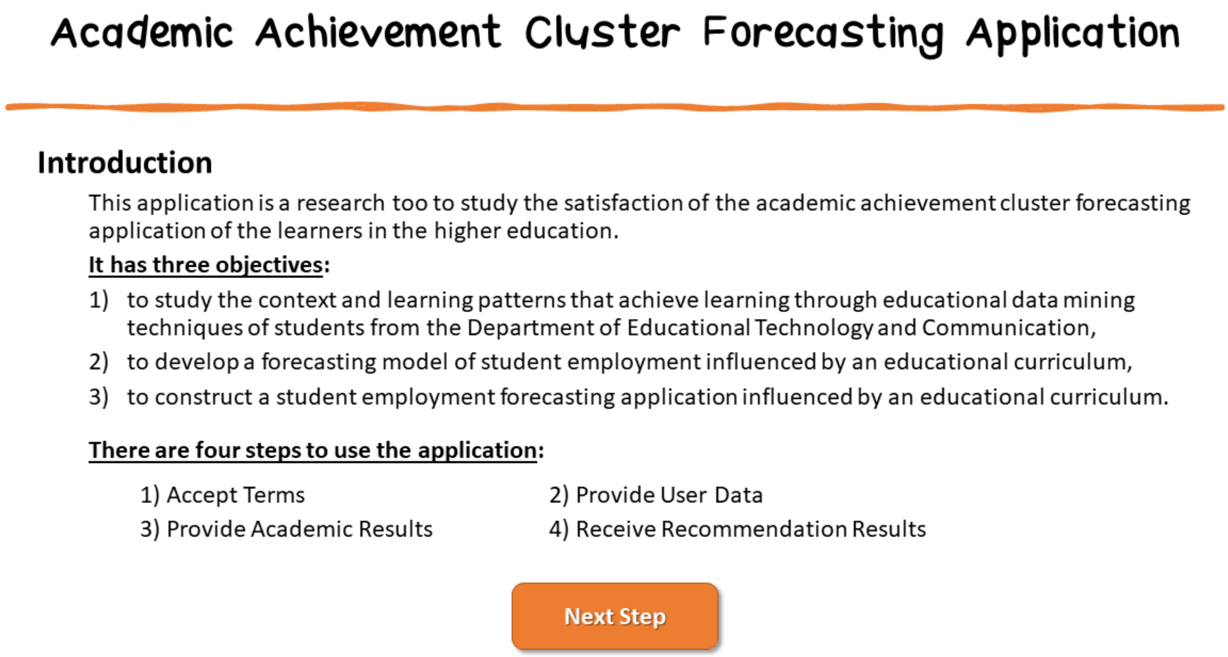

Fig. 5 Application Functionality and User Interface

Fig. 6 Application Functionality and User Interface

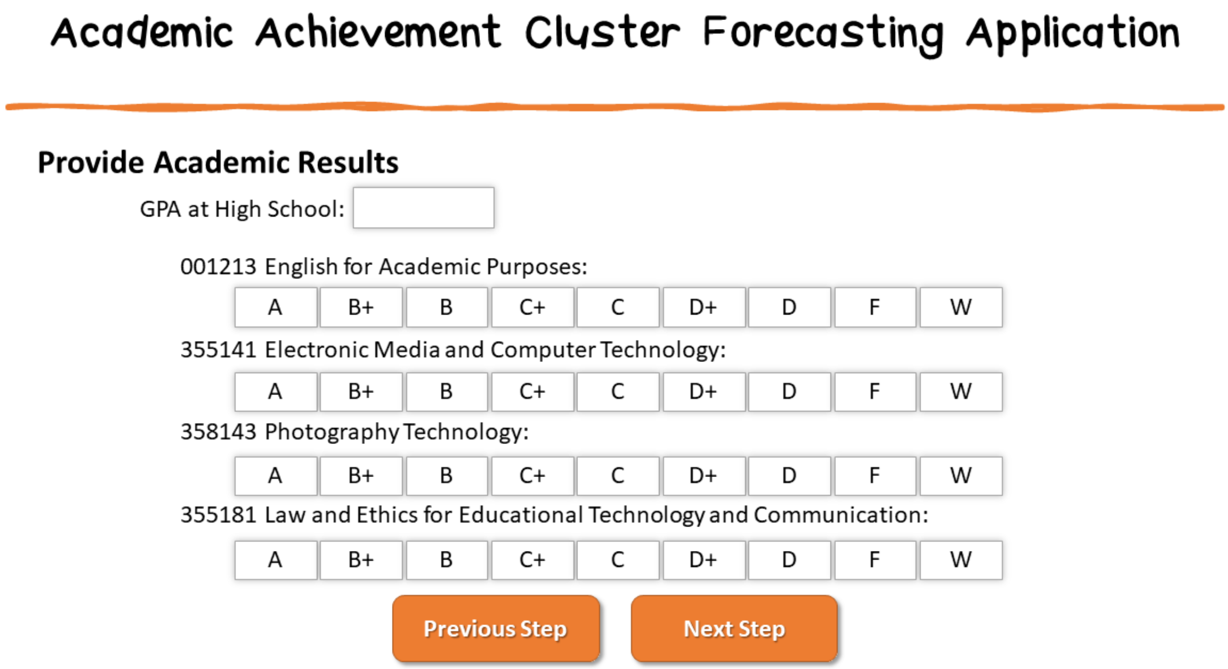

Fig. 7 Application Functionality and User Interface

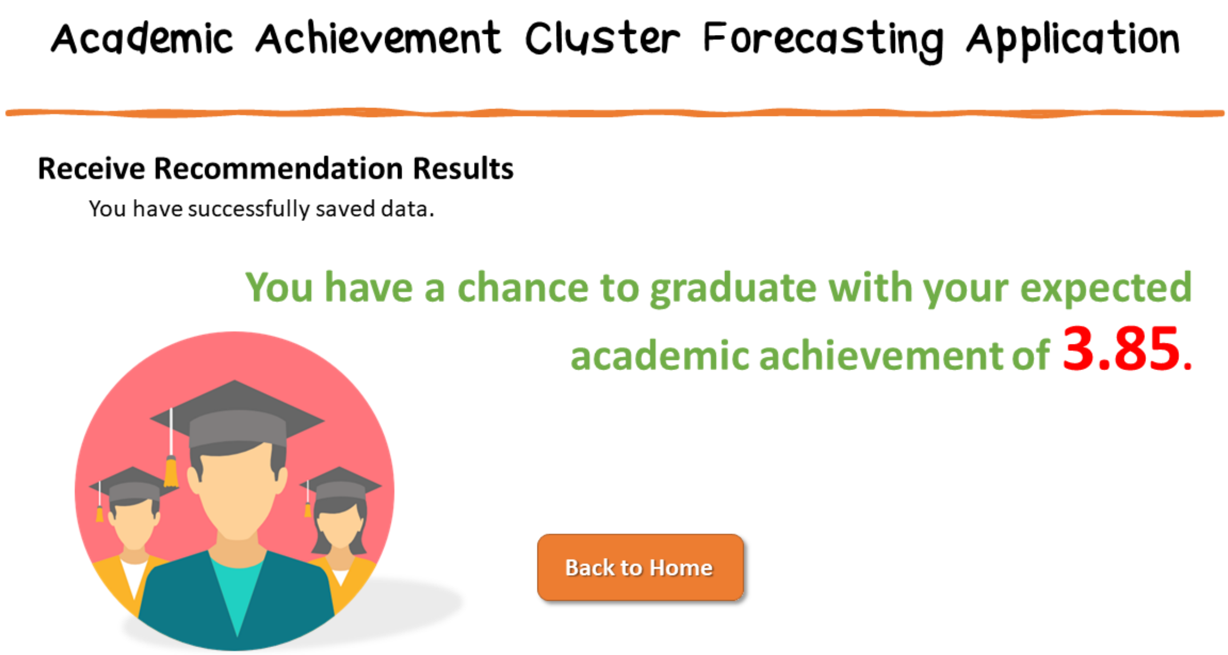

Fig. 8 Application Functionality and User Interface

Figure 5 to Figure 8 shows an academic achievement cluster forecasting application. It consists of four important steps. The first step introduces and clarifies the use of the application, as shown in Figure 5. The second and third steps are to fill in the required user information for the forecasting, as shown in Figure 6 and Figure 7. Finally, the application presents the forecast results based on user input, as shown in Figure 8.

After the application was developed, it was tested by a target audience of 30 samples presented in the next section.

4.2.5. Satisfaction with the Application

This section presents 30 samples of application satisfaction assessments from a specific target group in which analysis techniques were averaging and interpreting. The issues associated with assessing application satisfaction are presented in Table 2, while the satisfaction assessment results are presented in Table 14. In addition, scoring and interpretations were presented in the methodology section.

Table 14 shows the results of the satisfaction assessment of the application testing. It found that, overall, testers were highly satisfied with the application. An average value of 4.21 interpreted that a tester accepted the application.

Subsequently, the results will be gathered for discussion.

Table 14. Satisfaction with the Application

Section / State	Satisfaction with the Application		
	Mean	S.D.	Interpretation
Assessment of Functionality and Procedures			
State 1	4.02	0.69	Agree
State 2	4.25	0.68	Strongly Agree
Assessment of the System Performance			
State 3	4.24	0.65	Strongly Agree
State 4	4.23	0.62	Strongly Agree
State 5	4.35	0.63	Strongly Agree
State 6	4.32	0.60	Strongly Agree
Assessment of Convenience and UX-UI Design			
State 7	4.30	0.67	Strongly Agree
State 8	4.13	0.68	Agree
Assessment of the System Quality			
State 9	4.02	0.69	Agree
State 10	4.25	0.68	Strongly Agree
Overall	4.21	0.66	Strongly Agree

5. Discussion

The results discussion of this research is dedicated to the research objectives, which consisted of three key parts.

5.1. Discussions in the Context of Educational Program

In the research results in the context of the educational program, the researchers opened perspectives on two dimensions relevant to the educational program.

The first dimension is the students' achievement in the Bachelor of Art Program, Educational Technology, and Communications at the Faculty of Education, Naresuan University. The second dimension is the study of the context of learners' employment. It found that most students had a high level of academic achievement, as shown in Table 5. It also found that most students had higher academic achievement after graduation. It was an overall grade point average of 2.87 before enrolling in the program. After graduation, students had an overall grade point average of 3.13.

In contrast, the data on the employment status of post-graduate students are presented in Table 3. It found that the percentage of students being employed was only 63.33 percent. Additionally, only 41.11 percent of employed students work directly in their field.

The researchers' observations and doubts were that some learners graduated after the plan was designed, as shown in Table 4. The reason is the lack of competence in the educational program that students decided to register and study until graduation. Once the learners graduate, they seek and pursue desired careers that may not be consistent with their intended educational program. Since 33.33 percent of the students were not employed after graduation, the educational program should follow up with the learners after graduation.

5.2. Discussion in the Model and Application Development

The researchers develop models and applications to increase the small number of students in the low-achieving category, as shown in Table 5. It raises concerns that learners have a chance of failing to complete their studies. Therefore, the researchers developed models and applications for clustering student achievement forecasts. This tool was used to plan and monitor learners closely, as presented in the modeling and selection of the reasonable model in Figure 1 to Figure 4 and Table 6 to Table 13. Additionally, the researchers presented the application's user interface, as shown in Figures 5 to 8.

The researchers discussed two aspects. The first aspect is how to construct the academic achievement cluster. The second aspect is how to use the appropriate clusters to predict academic achievement. Regarding the process of constructing an achievement cluster, researchers used correlations of two data sets that consist of admission GPA and graduation GPA as defined and summarized in Table 5. The analytical tool for the development of rational achievement clusters uses two clustering techniques: the k-Means technique and the k-Medoids technique. The appropriate cluster numbers were determined using the elbow and k-determination techniques.

The conclusions obtained from these two techniques consist of the optimal cluster number of the k-Means technique, where k is equal to 7, as shown in Figure 1 and the k-Medoids technique, where k is equal to 8, as shown in Figure 2. The clarifications on the various indicators are presented in Table 6 to Table 11.

In terms of model development, the cluster results of both techniques (k-Means and k-Medoids) were used to develop a forecasting model based on five prediction techniques: decision tree, random forest, gradient boosted trees, decision stump, and random tree. The results showed that the gradient-boosted trees (GBT) model based on the datasets of k-Means clustering results provided the highest accuracy, as shown in Table 12. It was chosen for application development.

The discussion points out where all indicators of model performance testing from Table 12 and Table 13 are considered. It found that the overall performance of model testing was quite low in accuracy. The assumption that the researchers envisioned was a small number of datasets. 90 datasets were analyzed to develop the model. However, this is the actual data set collected from the educational program. Research should explore ways to predict higher academic achievement clusters. This issue has caught the attention of data science disciplines who want to perform deep analysis of small datasets to develop high-performance models for real datasets.

5.3. Discussion of the Satisfaction toward the Application

The application satisfaction assessment aims to study user attitudes to improve the application's suitability and compliance with the user's needs. The assessment criteria are based on the Likert Rating Scale, which consists of five levels defined in the "application satisfaction assessment section". Furthermore, assessment issues consist of four dimensions, which comprise 10 sub-items, as presented in Table 2.

The target group selected by the researchers to assess application satisfaction was a purposive sampling of 30 representatives. The application satisfaction assessment results are summarized in Table 14. It found that most of the representatives were satisfied with using the application. There was an overall average of 4.21 (S.D.=0.66) across all dimensions. It was considered a very high level of application acceptance and satisfaction. In addition, the representative issue offered the highest satisfaction at "State 5: Satisfaction with the user-friendly design and simple menu". There was an average satisfaction rating of 4.35 (S.D.=0.63). In addition, the issue with the lowest average satisfaction was "State 1: Satisfaction with the steps and processes of the system". There was an average satisfaction rating of 4.02 (S.D.=0.69).

As a preliminary review of this satisfaction assessment, researchers are satisfied with the assessment and believe that the application will benefit the learners and the educational program in the future. Suggestions can improve the process and adjust the descriptions in the Thai language for user convenience in the future.

6. Conclusion

The context of the student's employment is necessary and relevant to the context of the educational program. Therefore, developing a learning achievement forecasting model is necessary and significant in managing the educational program. In addition, this research identified goals and objectives in three important areas. The first objective is to study the context of student employment and academic achievement in the educational program. The second objective is to develop a learning achievement cluster forecasting model. The third objective is to assess the learning achievement cluster forecast application.

The research data were collected from 227 students at the Department of Educational Technology and Communication, the Faculty of Education, Naresuan University, during the academic year 2015 – 2020. The data used in this research were divided into two parts: data for developing the learning achievement cluster forecasting model and data for assessing the learning achievement cluster forecast application. In addition, the research instruments were divided into two categories. A first category is a tool for fundamental statistical analysis. It contains the frequency, mean, percentage, standard mean (S.D.), and interpretation. The second category is model development and model performance analysis tools. It contains the decision tree, random forest, gradient-boosted trees, decision stump, random tree, cross-validation methods, and confusion matrix performance.

The research findings are divided into three parts according to the objectives. The first part found that the overall students had a high level of grade point average (GPA). There was a 3.58 average overall student GPA. There were doubts about the percentage of student employment, where only 63.33% of students were employed. Moreover, only 41.11 percent of students were employed and met the educational program goals. In this regard, the educational program must review and find solutions. The second part is the development of the learning achievement cluster forecasting model. There are two steps: finding the correct number of learning achievement clusters and developing the learning achievement cluster forecasting model. Optimal cluster results showed that the k-Means technique combined with the Gradient Boosted Trees (GBT) technique offered the highest model efficiency. There was 71.11 percent accuracy.

The model is usable and acceptable; however, the model's performance is required to be further improvement. The last part is satisfaction with the evaluation of the application. It found that overall, representatives accepted the application with an overall satisfaction level of 4.21 (S.D.=0.66). Moreover, the researchers plan to test and deploy the application with those who are responsible for the educational program to create a strategy for managing performance and student achievement.

The success of this research reflects the value of data utilization, especially in the dimensions of the education industry. Nevertheless, Thailand still lacks data science personnel and technology, so it should be promoted and supported for further information access.

Acknowledgments

This research project was supported by the Faculty of Education at Naresuan University and the University of Phayao. In addition, this research was supported by many advisors, academicians, researchers, students, staff, and agencies. The authors would like to express their deepest gratitude for their support and collaboration in contributing to this research.

References

[1] N. Bunwirat and S. Boonsathorn,"Inquiry-Based Learning: an Effective Pedagogical Approach for Empowering 21st Century Learners and Education 4.0 in Thailand," *Political Science and Public Administration Journal*, vol. 9, no. 1, 2018.

[2] J. Siririn, "Higher Education 4.0: The Adaptation of Global University," *National Interest*, vol. 2, no. 6, pp. 49–55, 2021.

[3] T. Suksai, P. Suanpang, and R. Thangchitcharoenkhul, "A Digital Leadership Development Model for School Administrators in Basic Education to Fulfill the Thailand 4.0 Policy," *PSAKU International Journal of Interdisciplinary Research,* vol. 10, no. 2, 2021.

[4] S. Thonthong and S. Paengthai, "The Model of Early Childhood Education Management, Education era 4.0 of Educational Institutions under the Office of the Basic Education Commission," *Journal of Graduate MCU KhonKaen Campus*, vol. 6, no. 3, pp. 490–502, 2019.

[5] S. Wittayasin, "Education Challenges to Thailand 4.0," *International Journal on Integrated Education Development,* vol. 2, no. 2, pp. 29–35, 2018.

[6] K.B. Aviso, F.P.A. Demeterio, J.I.B. Janairo, R.I.G. Lucas, M.A.B. Promentilla, R.R. Tan and D.E.C. Yu, "What University Attributes Predict for Graduate Employability?," *Cleaner Engineering and Technology*, vol. 2, pp. 100069, 2021. Crossref, https://doi.org/10.1016/j.clet.2021.100069

[7] E. Courtial and C. Garrouste, "Model Predictive Control Strategy to Forecast Employability in Earth Sciences," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 10731–10736, 2014. Crossref, https://doi.org/10.3182/20140824-6-ZA-1003.02563

[8] K. K. C. Seah, J. Pan, and P. L. Tan, "Breadth of University Curriculum and Labor Market Outcomes," *Labour Economics,* vol. 65, pp. 101873, 2020. Crossref, https://doi.org/10.1016/j.labeco.2020.101873

[9] Ș. S. Nartgün and R. Özen, "Investigating Pedagogical Formation Students' Opinions about Ideal Teacher, Teaching Profession, Curriculum, Responsibility, Public Personnel Selection Examination (ppse) and Employment: A Metaphor Study," *Procedia - Social and Behavioral Sciences*, vol. 174, pp. 2674–2683, 2015. Crossref, https://doi.org/10.1016/j.sbspro.2015.01.951

[10] L. Gao, Z. Zhao, C. Li, J. Zhao, and Q. Zeng, "Deep Cognitive Diagnosis Model for Predicting Students' Performance," *Future Generation Computer Systems*, vol. 126, pp. 252–262, 2022. Crossref, https://doi.org/10.1016/j.future.2021.08.019

[11] P. Nuankaew, “Dropout Situation of Business Computer Students, University of Phayao,” *International Journal of Emerging Technologies in Learning (iJET),* vol. 14, no. 19, 2019. Crossref, https://doi.org/10.3991/ijet.v14i19.11177

[12] A. Konstanti and A. Moropoulou, "Hybrid Educational Methodology for the Cognitive Domain of Built Heritage Protection Interconnecting Secondary with Tertiary Level Education," *International Journal of Engineering Pedagogy (iJEP)*, vol. 3, no. 4, 2013. Crossref, https://doi.org/10.3991/ijep.v3i4.2761

[13] E. Araka, E. Maina, R. Gitonga, and R. Oboko, "A Conceptual Model for Measuring and Supporting Self-Regulated Learning using Educational Data Mining on Learning Management Systems," in *2019 IST-Africa Week Conference (IST-Africa)*, pp. 1–11, 2019. Crossref, https://doi.org/10.23919/ISTAFRICA.2019.8764852

[14] W. S. Nuankaew, P. Nuankaew, D. Teeraputon, K. Phanniphong, and S. Bussaman, “Perception and Attitude Toward Self-Regulated Learning of Thailand’s Students in Educational Data Mining Perspective,” *International Journal of Emerging Technologies in Learning (iJET),* vol. 14, no. 9, pp. 34–49, 2019. Crossref, https://doi.org/10.3991/ijet.v14i09.10048

[15] C. Romero and S. Ventura, "Educational Data Mining and Learning Analytics: An Updated Survey," *WIREs Data Mining and Knowledge Discovery*, 2020. Crossref, https://doi.org/10.1002/widm.1355

[16] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," *IEEE Transactions on Systems, Man, and Cybernetics, Part C Applications and Reviews*, vol. 40, no. 6, pp. 601–618, 2010. Crossref, https://doi.org/10.1109/TSMCC.2010.2053532

[17] P. Nuankaew and W. S. Nuankaew, "Student Performance Prediction Model for Predicting Academic Achievement of High School Students," *European Journal of Educational Research*, vol. 11, no. 2, pp. 949–963, 2022. Crossref, https://doi.org/10.12973/eu-jer.11.2.949

[18] M. Sarfraz, K. F. Khawaja, and L. Ivascu, "Factors Affecting Business School Students' Performance During the COVID-19 Pandemic: A Moderated and Mediated Model," *The International Journal of Management Education*, vol. 20, no. 2, pp. 100630, 2022. Crossref, https://doi.org/10.1016/j.ijme.2022.100630

[19] R. Mitola, E. Rinto, and E. Pattni, “Student Employment as a High-Impact Practice in Academic Libraries: A Systematic Review,” *The Journal of Academic Librarianship*, vol. 44, no. 3, pp. 352–373, 2018. Crossref, https://doi.org/10.1016/j.acalib.2018.03.005

[20] A. Kenny, V. Dickson-Swift, N. DeVecchi, C. Phillips, B. Hodge, and Y. Masood, "Evaluation of a Rural Undergraduate Nursing Student Employment Model," *Collegian*, vol. 28, no. 2, pp. 197–205, 2021. Crossref, https://doi.org/10.1016/j.colegn.2020.07.003

[21] B. S. Baert, B. Neyt, T. Siedler, I. Tobback, and D. Verhaest, "Student Internships and Employment Opportunities After Graduation: A Field Experiment," *Economics of Education Review*, vol. 83, pp. 102141, 2021. Crossref, https://doi.org/10.1016/j.econedurev.2021.102141

[22] J. Mitchell, "Juggling Employment and Studies: Nursing Students' Perceptions of the Influence of Paid Employment on their Success," *Nurse Education Today*, vol. 92, pp. 104429, 2020. Crossref, https://doi.org/10.1016/j.nedt.2020.104429

[23] K. R. B. Calma, A. Williams, S. McInnes, and E. Halcomb, "New Graduate Employment in General Practice: Perceptions of Final-Year Nursing Students," *Nurse Education in Practice*, vol. 54, pp. 103115, 2021. Crossref, https://doi.org/10.1016/j.nepr.2021.103115

[24] P. Bryła, "The Impact of International Student Mobility on Subsequent Employment and Professional Career: A Large-scale Survey Among Polish Former Erasmus Students," *Procedia - Social and Behavioral Sciences*, vol. 176, pp. 633–641, 2015. Crossref, https://doi.org/10.1016/j.sbspro.2015.01.521

[25] E. Hartikainen, J. Suhonen, and E. Sutinen, "IMPDET - an Online PhD Study Program in Educational Technology," in *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*, pp. 578–580, 2006. Crossref, https://doi.org/10.1109/ICALT.2006.1652508

[26] P. Bakker, D. Carrington, A. Goodchild, I. Hayes, H. Purchase, and P. Strooper, "The Communicating Technologist: an Educational Challenge," in *Proceedings Frontiers in Education 1995 25th Annual Conference. Engineering Education for the 21st Century*, vol. 2, pp. 4a4.1-4a4.4, 1995. Crossref, https://doi.org/10.1109/FIE.1995.483173

[27] R. Hartley, Kinshuk, and R. Koper, "Preparing Coming Generations of Educational Technologists: Technology, Pedagogy and Curricula–Changes and Challenges," in *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*, pp. 890–892, 2007. Crossref, https://doi.org/10.1109/ICALT.2007.208

[28] Haryono, Y. Utanto, Budiyono, E. Subkhan, and S. Zulfikasari, “The Implementation of Educational Technologists’ Competencies in Improving Learning Quality,” in *2019 5th International Conference on Education and Technology (ICET)*, pp. 76–80, 2019. Crossref, https://doi.org/10.1109/ICET48172.2019.8987215

[29] C. Schröer, F. Kruse, and J. M. Gómez, “A Systematic Literature Review on Applying CRISP-DM Process Model,” *Procedia Computer Science*, vol. 181, pp. 526–534, 2021. Crossref, https://doi.org/10.1016/j.procs.2021.01.199

[30] M. Cazacu and E. Titan, "Adapting CRISP-DM for Social Sciences," *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, vol. 11, no. 2Sup1, 2021. Crossref, https://doi.org/10.18662/brain/11.2Sup1/97

[31] P. Nuankaew and P. Temdee, "Matching of Compatible Different Attributes for Compatibility of Members and Groups," *International Journal of Mobile Learning and Organisation*, vol. 13, no. 1, pp. 4–29, 2019. Crossref, https://doi.org/10.1504/IJMLO.2019.096